

# Semantic Web

## Evolution oder Revolution eines Mediums?

- Informatikunterricht der Zukunft
- 2. März 2006
- Dr. Harald Sack
- Friedrich-Schiller-Universität Jena

# Semantic Web

## Evolution oder Revolution eines Mediums?

More machine processable than before

only another  
New, improved  
**SEMANTIC**  
Web  
defining meaning

**HYPE...!?**

May be incompatible with existing XML tools. Databases may take up to ten times as much memory and 24 hours to load.

# Semantic Web

## Evolution oder Revolution eines Mediums?

### Inhalte:

1. Das World Wide Web – 1990 bis heute
2. Die Grenzen des WWW
3. Die Vision des Semantic Web
4. Wie das Semantic Web die Welt verändert

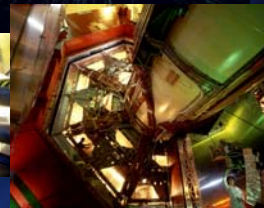
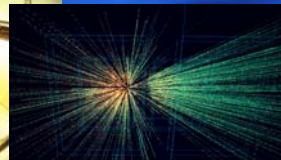


## 1. Das World Wide Web – 1990 bis heute

### ● Die Geburt des WWW



Robert Cailliau Tim Berners-Lee



Das **WWW** wurde 1990 am europäischen Kernforschungszentrum CERN geboren

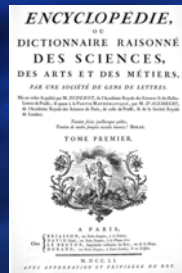
# 1. Das World Wide Web – 1990 bis heute

## ● Die Geburt des WWW

- Aber die ursprüngliche Idee reicht viel weiter zurück....



Agostino Ramelli (1588),  
Le diverse et artificieuse machine;  
composte in lingua Italiana et Francese



Encyclopédie ou Dictionnaire raisonné  
des sciences, des arts et des métiers  
(1772)



Denis Diderot  
(1713-1784)



Jean-Baptiste le Rond  
d'Alembert  
(1717-1783)

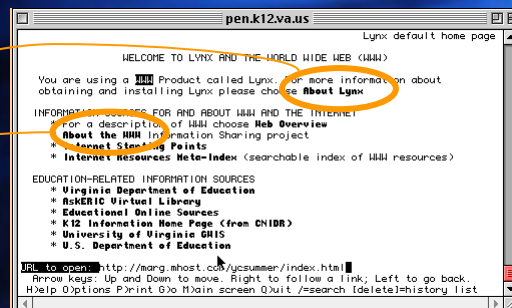
Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

5

# 1. Das World Wide Web – 1990 bis heute

## ● Die Erfolgstory des WWW

- Der Browser als universale und intuitive Benutzerschnittstelle



Lynx 1993

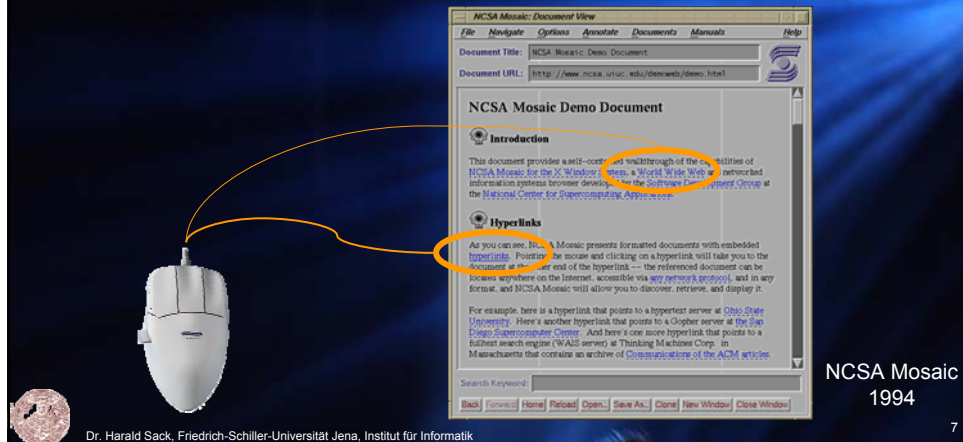
Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

6

# 1. Das World Wide Web – 1990 bis heute

- Die Erfolgstory des WWW

- Der Browser als universale und intuitive Benutzerschnittstelle



NCSA Mosaic  
1994

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

7

# 1. Das World Wide Web – 1990 bis heute

- Die Erfolgstory des WWW

- Der Browser als universale und intuitive Benutzerschnittstelle



Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

8



# 1. Das World Wide Web – 1990 bis heute

## ● Das WWW heute – Web 2.0

### ○ Privatkommunikation, Broadcast, Interaktivität

- **Web 1.0** orientierte sich an der Funktionsweise der klassischen Massenmedien → **Broadcast, Einweg-Kommunikation**
- Nutzer ist von den übrigen Teilnehmern des WWW abgekapselt
- Nutzer als anonyme Masse, die **lesen, kaufen** und **Werbebanner anklicken** sollen

Zu Hause günstig mit dem Handy telefonieren:



Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

9

# 1. Das World Wide Web – 1990 bis heute

## ● Das WWW heute – Web 2.0

### ○ Privatkommunikation, Broadcast, Interaktivität

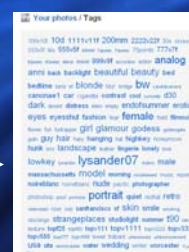
- **Web 2.0** bietet dem Nutzer die Freiheit, Inhalte aktiv mitzugestalten



WebLogs  
(Blogs)



Flickr



Tag-Clouds  
Folksonomies

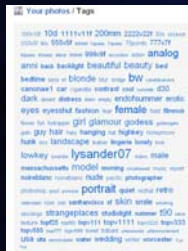
Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

10

# 1. Das World Wide Web – 1990 bis heute

## • Das WWW heute – Web 2.0

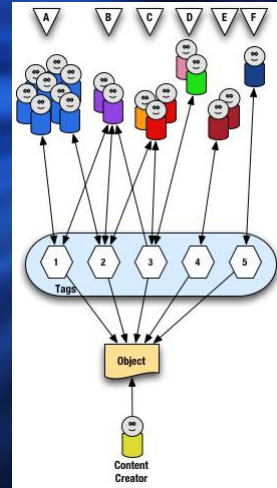
- Web 2.0 bietet dem Nutzer die Freiheit, Inhalte aktiv mitzugestalten



Tag-Clouds  
Folksonomies

Nutzer kennzeichnen und katalogisieren Web-Inhalte mit eigenen Schlagworten (Tags)

„Soziale Netze“



# 1. Das World Wide Web – 1990 bis heute

## • Das WWW heute – Web 2.0

- Privatkommunikation, Broadcast, Interaktivität
  - Desktop-Anwendungen migrieren zu Web-Anwendungen (Webtop)



GMail  
E-Mail-Client



S5  
Präsentationssystem



writeley  
Team-Textverarbeitung

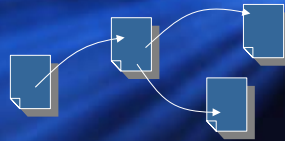


Wikipedia  
online Lexikon

# 1. Das World Wide Web – 1990 bis heute

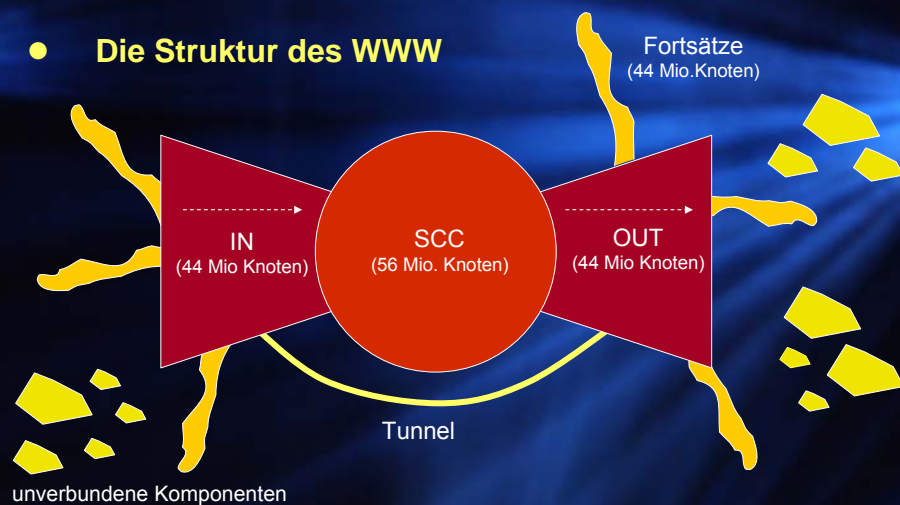
- **Das WWW heute - (Stand: 03/2006)**

- Das **World Wide Web** bietet Zugriff auf eine gigantisches „**Informationsuniversum**“
- Schätzungen gehen von **mehr als 55 Milliarden** Dokumenten im WWW aus
- das WWW ist ein **Graph**
- **Konnektivität des WWW**
  - **Schwach verbundene Komponenten**
    - Hyperlinks werden als **ungerichtete Kanten** interpretiert
    - einzelne Komponente mit ca. **90%** aller untersuchten Dokumente
  - **Stark verbundene Komponenten**
    - Hyperlinks werden als **gerichtete Kanten** interpretiert
    - **28%** der untersuchten Dokumente bilden eine stark verbundene **Kernkomponente (SCC)**



# 1. Das World Wide Web – 1990 bis heute

- **Die Struktur des WWW**



unverbundene Komponenten

Broder e.a.: Graph structure in the web, WWW9, 2000

## 2. Die Grenzen des World Wide Web

### ● Informationsuniversum WWW

- etwa **20 Mrd. Dokumente** des WWW sind in Suchmaschinen wie z.B. Google **indiziert** (Stand 11/2005)
- Dokumentenbestand im WWW **doppelt sich** **alle 6 Monate**
- ein Ende dieser Entwicklung **schon nicht absehbar**

### ○ Probleme

- Wie finde ich die richtigen (wichtigen) Informationen (vollständig)?
- Wie sorge ich dafür, dass ich die von mir angebotenen Informationen **schonlich** findet?
- Kann die Informationssuche (zuverlässig) automatisiert werden?
- Kann ich der angebotenen Information auch tatsächlich trauen?

## 2. Die Grenzen des World Wide Web

### ● Das Ende des WWW?

- Was nutzt uns die ganze Informationsfülle?



- was ist **wichtig**, was nicht?
- was ist **Information**, was ist **Werbung**?
- was **bedeutet** die Information?
- wie **glaubhaft** ist die Information?
- was **gehört zusammen**?
- was ist **redundant**?



## 2. Die Grenzen des World Wide Web

### ● Das Ende des WWW?

- Was nutzt uns die ganze Informationsfülle?



Mensch verfügt über **Kontextwissen / Weltwissen** und kann daher Information (meist) richtig deuten...

## 2. Die Grenzen des World Wide Web

### ● Das Ende des WWW?

- Was nutzt uns die ganze Informationsfülle?
- → Das WWW ist zur **Nutzung durch den Menschen** bestimmt
- WWW basiert auf der **Markupsprache HTML**
- HTML beschreibt
  - **wie** Informationen **dargestellt** werden sollen,
  - **wie** Informationen miteinander **verknüpft** werden können,
  - aber **nicht, was diese Informationen bedeuten....**



Semantik ???

## 2. Die Grenzen des World Wide Web

### ● Semantik

Die **Semantik (Bedeutungslehre)** ist das Teilgebiet der Sprachwissenschaft (*Linguistik*), das sich mit

- ❑ Sinn und
- ❑ Bedeutung

von **Sprache** beziehungsweise **sprachlichen Zeichen** befasst.

Die Semantik kümmert sich um die Frage, wie **Sinn und Bedeutung von komplexen Begriffen** aus denen von **einfachen Begriffen** abgeleitet werden können und stützt sich dabei in der Regel auf die **Syntax**.

## 2. Die Grenzen des World Wide Web

### ● Das Ende des WWW?

- → Das WWW ist zur **Nutzung durch den Menschen** bestimmt



Programm (z.B. Robot) kann nicht unterscheiden zwischen wichtiger / unwichtiger Information

## 2. Die Grenzen des World Wide Web

### • Das Ende des WWW?

- → Das WWW ist zur **Nutzung durch den Menschen** bestimmt
- **Problemfelder:**
  - [1] **Informationssuche**  
→ „wo finde ich...?“
  - [2] **Informationsextraktion**  
→ „was ist wichtig?“
  - [3] **Wartung**  
→ „sind meine Daten (noch) konsistent?“
  - [4] **Personalisierung**  
→ „was mich interessiert...“

## 2. Die Grenzen des World Wide Web

### • Problemfeld 1: Informationssuche



#### • Keyword-Suche führt zu **vielen nicht relevanten Ergebnissen**

- Bedeutungsunterschiede
- Homonyme
- unterschiedlicher Kontext

#### • Keyword-Suche findet **nicht alle relevanten Ergebnisse**

- Synonyme
- fehlende Kontext-präzisierung



**SEMANTISCHE  
Verschlagwortung**



## 2. Die Grenzen des World Wide Web

### ● Problemfeld 2: Informationsextraktion



- kann nur von menschlichem "Agenten" **korrekt** durchgeführt werden

- Software-Agent verfügt nicht über
  - **(kulturelles) Kontextwissen**
  - **Weltwissen**
- um Informationen aus der Text- / Bilddarstellung zu lösen



**Einbindung zusätzlicher SEMANTIK**

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

23

## 2. Die Grenzen des World Wide Web

### ● Problemfeld 3: Wartung



- je komplexer und größer die WebSite, desto aufwändiger die Wartung der oft nur schwach-strukturierten Daten

#### • Problem:

- Konsistenz
- Korrektheit
- Aktualität



**maschinelle Wartung benötigt Information über SEMANTIK**

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

24



## 2. Die Grenzen des World Wide Web

### ● Problemfeld 4: Personalisierung



- Anpassung des dargestellten Informationsinhalts an die **persönlichen Bedürfnisse und Anforderungen**

- **Problem:**  
„woher kommen die benötigten Informationen?“



erfordert Möglichkeit der **Repräsentation von SEMANTIK**

- Informationsanbieter vs. Informationskonsument

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

25

## 3. Die Vision des Semantic Web

### ● Die ursprüngliche Vision des WWW

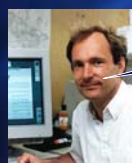
*I know what you mean...*



„The Web was designed as an information space, with the goal that it should be useful **not only for human-human communication, but also that machines would be able to participate and help...**“

Tim Berners-Lee, Semantic Web Roadmap, Sept 1998

**HTML** als Sprache zur Informationsdarstellung im WWW besitzt **keine Möglichkeit, die Bedeutung** der dargestellten Information **auszudrücken**



Tim Berners-Lee

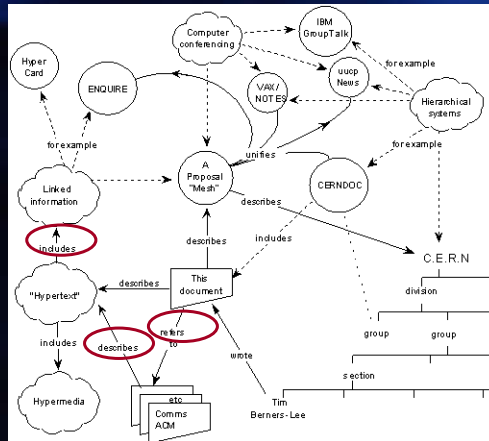
*Das volle Potenzial des WWW wird nicht genutzt...*

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

26

### 3. Die Vision des Semantic Web

#### • Die ursprüngliche Vision des WWW



• ursprüngliche Beschreibung des WWW **beschreibt unterschiedliche Beziehungen zwischen Dokumenten**

- describes
- includes
- refers to
- ...

• diese Beziehungen werden heute im WWW als nicht-attribuiertes Hyperlink ausgedrückt

→ Semantik ging verloren!

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

27

### 3. Die Vision des Semantic Web

#### • Die Vision des Semantic Web

- 1997, WWW7 in Brisbane, Australien
  - TBL stellt seine Vision eines „Global Reasoning Web“ vor
- 1998, WWW8 in Toronto, Kanada
  - TBL präsentiert seine „**Semantic Web Roadmap**“
- *Was ist das Semantic Web?*
- *Was kann das Semantic Web?*
- *Was ist das Semantic Web nicht?*
- *Wird das Semantic Web unser Leben verändern?*
- ...



Tim Berners-Lee

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

28

### 3. Die Vision des Semantic Web

- Was ist das Semantic Web?

- „The Semantic Web is a **web of data**, in some ways like a **global database**.“
- „...a **universal Web of semantic assertions**“
- „...emphasises the **intelligent manipulation of the knowledge** that is made available via the web.“
- „The Semantic Web is an **extension of the current web** in which **information is given well-defined meaning**, better enabling **computers and people** to work in cooperation“

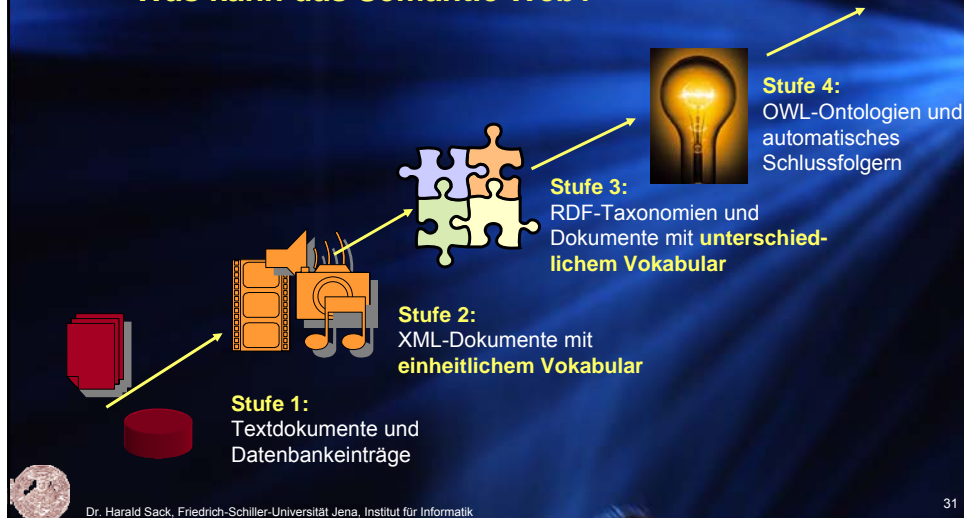
### 3. Die Vision des Semantic Web

- Was ist das Semantic Web?

- Das Semantic Web ist **eine Art globaler Datenbank**, das ein **universales Netz semantischer Aussagen** bereit hält
- Das Semantic Web kann sein Potenzial nur dann ausschöpfen, wenn **Daten** sowohl von **Programmen als auch von Menschen ausgetauscht und verarbeitet** werden können
- **Programme** müssen heterogene Daten austauschen und gemeinsam verarbeiten können, **auch wenn diese Programme völlig unabhängig voneinander entwickelt wurden**.

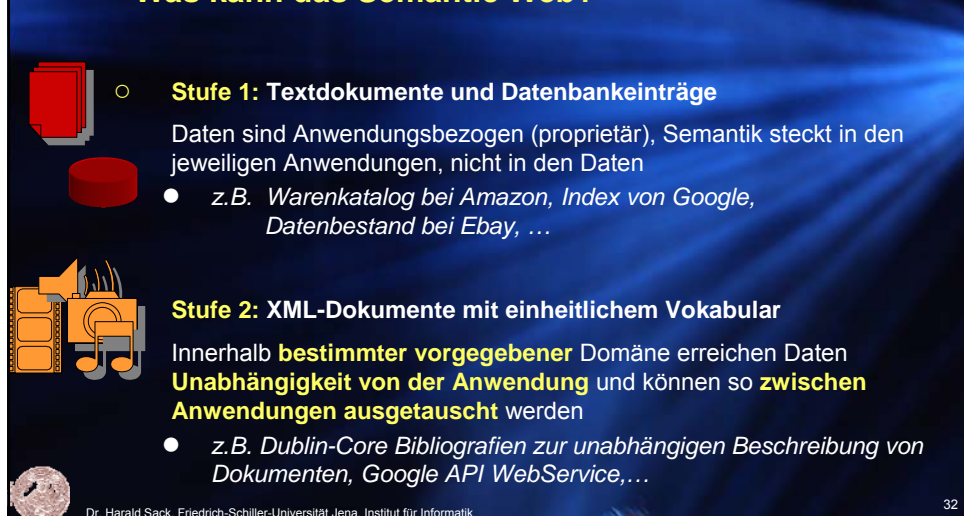
### 3. Die Vision des Semantic Web

- Was kann das Semantic Web?



### 3. Die Vision des Semantic Web

- Was kann das Semantic Web?





### 3. Die Vision des Semantic Web

- Was kann das Semantic Web?



- Stufe 3: RDF-Taxonomien und Dokumente mit unterschiedlichem Vokabular

Daten stammen aus **unterschiedlichen Domänen** und werden durch hierarchisch strukturierte **Taxonomien** klassifiziert. Einfache Beziehungen zwischen den einzelnen Kategorien dienen der Kombination von Daten.

- z.B. „Webtechnologien“ ist eine „Vorlesung“ ist eine „Lehrveranstaltung“
- z.B. „Theoretische Informatik“ gehört zum „Institut für Informatik“ gehört zur „Fakultät Mathematik und Informatik“ gehört zur „FSU Jena“



### 3. Die Vision des Semantic Web

- Was kann das Semantic Web?



- Stufe 4: OWL-Ontologien und automatisches Schlussfolgern

- **Neue Daten** können aus bestehenden Daten gewonnen werden durch die Anwendung von **Schlussfolgerungsregeln**. **Daten, Eigenschaften von Daten und Beziehungen zwischen Daten** lassen sich semantisch beschreiben.

- z.B. automatisches Übersetzen eines Dokuments von einer Domäne in eine andere
- z.B. Formulieren komplexer Abhängigkeiten und Bedingungen
- → Anwendung „versteht“ Daten



### 3. Die Vision des Semantic Web

- **Was kann das Semantic Web?**

- Kann man den semantischen Daten auch trauen?

- **Authentizität**  
Ist der mutmaßliche Urheber (Autor) auch der Tatsächliche?
- **Integrität**  
Wurden die Daten manipuliert?

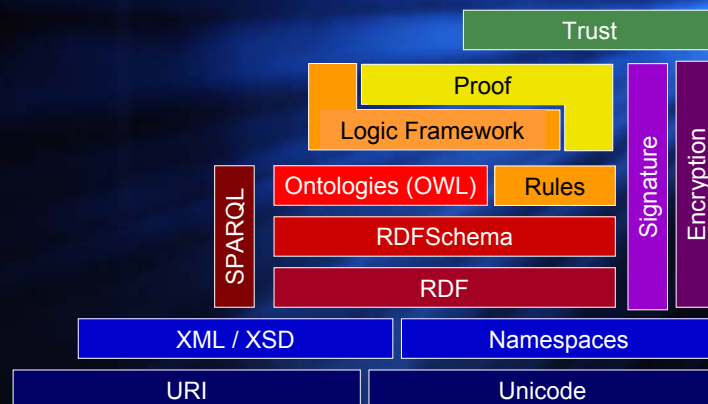


- **Public Key Kryptografie und digitale Signaturen**

- Semantische Daten müssen signiert werden, damit Inferenzsysteme deren Zuverlässigkeit mit in Betracht ziehen können
- Semantisches Dokument = Sammlung von Aussagen + zugehörige Signaturen

### 3. Die Vision des Semantic Web

- **Was kann das Semantic Web?**



### 3. Die Vision des Semantic Web

- **Was ist das Semantic Web nicht?**
  - **Nicht Teil des Semantic Webs sind (jetzt und in naher Zukunft)**
    - assoziatives Schlussfolgern
    - räumliches Denken
    - Erkennung von Text, Bildern, Gesten, etc...
    - Komplexe Entscheidungsprozeduren
    - Schlussfolgern aus unvollständigen oder kontextabhängigen Informationen
    - ...
  - „Just as Prolog is not AI but merely a useful tool for it, SW might be just a good tool for AI“

### 4. Wie das Semantic Web die Welt verändert

- **Suchmaschinen**
  - *Eingabe in Suchmaschine:* **“GOLF von 2000“**
  - **Nachfrage zu “GOLF“:**  
PKW oder SPORTART oder SONSTIGES ?  
Verfeinerung: **PKW**
  - **Nachfrage zu “2000“:**  
Baujahr oder technische Parameter?  
Verfeinerung: **Baujahr**
  - **Nachfrage zu “GOLF von 2000“:**  
Verkaufsangebote, Dokumentation oder sonstige Artikel ?  
Verfeinerung: **Dokumentationen**
  - **Nachfrage zu “Dokumentationen“:** ...



## 4. Wie das Semantic Web die Welt verändert

### ● Suchmaschinen (II)

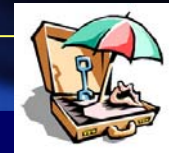
- Eingabe in Suchmaschine: **"GOLF von 2000"**
- Sonstige (persönliche) verfügbare Informationen:
  - Benutzer hat in den vergangenen Monaten regelmäßig die Ergebnisse von Profi-Golfturnieren abgerufen
  - Benutzer hat im vergangenen Jahr eine Garnitur Golfschläger (via Online-Shopping) gekauft
  - Benutzer hat die RSS-Feeds zu Golf-Fachpublikationen im Web abonniert
  - Benutzer hat den Bildschirm-Hintergrund „Pebble-Beach“ gewählt
  - Benutzer besitzt die Rechte an der Domain **www.playing-golf.de**
  - ...



## 4. Wie das Semantic Web die Welt verändert

### ● Reiseplanung

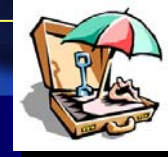
- Eingabe an den persönlichen Agenten:  
**Urlaub, mind. 10 Tage, mit Familie, am Meer < 2000€**
- Agent überprüft **Terminkalender** von **jedem Familienmitglied** und identifiziert entsprechende Freiräume
- Agent überprüft **Urlaubsangebote** innerhalb der gefundenen Termine entsprechend den Anforderungen
  - Orte einschränken
  - Familienfreundlichkeit?
  - Gibt es in der Nähe einen Golfplatz?
  - ...
- Agent erfragt **freie Kapazitäten** und stellt eine Auswahl der ermittelten Angebote entsprechend bekannter Nutzerpräferenzen zusammen





## 4. Wie das Semantic Web die Welt verändert

### ● Reiseplanung (II)



- *Eingabe an den Agenten:* **“Buche Reise xy”**
- Agent führt notwendige **Buchungen** durch
  - Fluggesellschaft / Bahn / ...
  - Hotel
  - Mietwagen (der notwendigen Größe...)
  - Ausflüge vor Ort
  - ...
- Agent veranlasst **Finanztransaktionen**
- Agent trägt Termin in die **Kalender** der Familienmitglieder ein
- Agent erstellt **Routenplanung** für das Navigationsgerät im familieneigenen PKW für die Fahrt zum Flughafen und reserviert einen günstigen Platz im Flughafenparkhaus

## Semantic Web

### Evolution oder Revolution eines Mediums?

### ● Zusammenfassung

- Das **Semantic Web** soll es (autonom agierenden) Anwendungsprogrammen ermöglichen
  - heterogene, über das Web zugreifbare Daten (**Web Data + Web Services**) miteinander in Beziehung zu setzen,
  - diese Daten gemeinsam zu nutzen (**Synergieeffekte**),
  - und aus diesen neue Informationen zu gewinnen (**Inferenzmechanismen**)
- Semantic Web und WWW ergänzen sich gegenseitig

# Semantic Web

Evolution oder Revolution eines Mediums?

More machine processable than before

More machine processable than before

SEMANTIC

May be incompatible with existing XML tools. Databases may take up to ten times as much memory and 24 hours to load.

beware of the HYPE...

Dr. Harald Sack, Friedrich-Schiller-Universität Jena, Institut für Informatik

43